Akademie věd České republiky
Ústav teorie informace a automatizace, v.v.i.

Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation

RESEARCH REPORT

Radomír Žemlička, Kamil Dedecius

Distributed Sequential Zero-Inflated Poisson Regression

No. 2393 December 7, 2021

ÚTIA AV ČR, P.O.Box 18, 182 08 Prague, Czech Republic
Tel: +420 266052570, Fax: +420 266052068, Url: http://www.utia.cas.cz,
E-mail: dedecius@utia.cas.cz

Abstract

The zero-inflated Poisson regression model is a generalized linear model (GLM) for non-negative count variables with an excessive number of zeros. This letter proposes its low-cost distributed sequential inference from streaming data in networks with information diffusion. The model is viewed as a probabilistic mixture of a Poisson and a zero-located Dirac component, whose probabilities are estimated using a quasi-Bayesian procedure. The regression coefficients are inferred by means of a weighted Bayesian update. The network nodes share their posterior distributions using the diffusion protocol.

1 Introduction

Fully distributed *sequential* (online) modeling of various stochastic processes from streaming data in networks with information diffusion is an established discipline in the signal processing domain [2, 1, 3]. Naturally, the main focus has been given to continuous variables that abound in practice. This originally gave rise to the diffusion recursive least squares [4], least mean squares [5, 6, 7], or Kalman filters [8, 9], followed by many variants. The continuous data models are frequently used for discrete counts data and work well if the values are high enough. However, for low counts and/or excessive zeros they are doomed to fail and the use of the generalized linear models (GLMs) is inevitable [10, 11].

In order to reflect the popularity of count models in epidemiology, finance, transportation, physics, or networking [12], a collaborative method for sequential (online) inference of the Poisson regression model was recently proposed by the authors [13]. The present paper extends it to the zero-inflated cases, where the excessive number of zeros cannot be attributed to the Poisson distribution [14]. There are several novel aspects. First, the problem is formulated as a mixture estimation task, where the excessive zeros are due to a Dirac distribution located at 0. The sequential estimation exploits the quasi-Bayesian idea [15], proposed originally for the linear regression models. Next, the inference of the Poisson component is stabilized for zero measurements. Finally, a procedure for merging the posterior estimates in a network of collaborating agents is proposed.

2 The ZIP Model

Let us consider a *sequential* (online) discrete-time modeling of a streaming stochastic process $\{Y_t; t = 1, 2, \ldots\}$ with observations $y_t \in \mathbb{N}_0$ that are mutually independent. The variables Y_t are determined by the regressors $x_t \in \mathbb{R}^n$, and a vector of possibly slowly time-varying unknown regression coefficients $\beta_t \in \mathbb{R}^n$. The link between the linear predictor $\beta_t^\mathsf{T} x_t$ and observations y_t characterizes the GLM [10]

$$\mathbb{E}[Y_t|x_t, \beta_t] = \exp(\beta_t^\mathsf{T} x_t), \tag{1}$$

where $\exp(\cdot)$ – the inverse link function – provides the Poisson regression [10, 11]

$$Y_t|x_t, \beta_t \sim \mathcal{P}\left(\exp(\beta_t^\mathsf{T} x_t)\right). \tag{2}$$

Recently, we formulated a Bayesian procedure for its sequential estimation from streaming data [13]. However, in many applications the counts tend to include significantly more zeros

than expected by the Poisson distribution [12]. In this case we need to adopt the zero-inflated Poisson distribution (ZIP) [14],

$$Y_t \sim ZIP\left(\exp(\beta_t^\intercal x_t), \phi_t\right), \quad \phi_t \in [0,1], \tag{3}$$

whose probability density function is expressed [14]

$$f(y_t|x_t,\beta_t,\phi_t)=\begin{cases}\phi_t+(1-\phi_t)\exp\left(-\exp(\beta_t^\intercal x_t)\right) & \text{if } y_t=0,\\ (1-\phi_t)\frac{\exp(\beta_t^\intercal x_t y_t-\exp(\beta_t^\intercal x_t))}{y_t!} & \text{otherwise.}\end{cases}$$

Apparently, it is possible to represent the ZIP distribution (3) by a convex combination (probabilistic mixture) of the Dirac distribution located at the origin, and a Poisson distribution,

$$Y_t|x_t,\beta_t,\phi_t \sim \phi_t \mathcal{D}irac(0) + (1-\phi_t)\mathcal{P}\left(\exp(\beta_t^\intercal x_t)\right). \tag{4}$$

The probability density then reads

$$f(y_t|x_t,\beta_t,\phi_t) = \phi_t\,\delta_0 + (1-\phi_t) f_{\mathcal{P}}(y_t|x_t,\beta_t), \tag{5}$$

where δ_0 is the Dirac measure at 0, and $f_{\mathcal{P}}(\cdot)$ is the Poisson density

$$f_{\mathcal{P}}(y_t|x_t,\beta_t) = \frac{\exp(\beta_t^\intercal x_t y_t - \exp(\beta_t^\intercal x_t))}{y_t!}. \tag{6}$$

The subscripts $\mathcal{D}$ and $\mathcal{P}$ will denote the respective distributions in the sequel. The expected value and the variance are

$$\mathbb{E}[Y_t|x_t,\beta_t,\phi_t] = (1-\phi_t)\exp(\beta_t^\intercal x_t), \tag{7}$$

$$\mathrm{var}(Y_t|x_t,\beta_t,\phi_t) = (1-\phi_t)\left[\exp(\beta_t^\intercal x_t) + \phi_t \exp(2\beta_t^\intercal x_t)\right]. \tag{8}$$

3 Sequential Estimation

The ZIP regression model inference faces two major difficulties. First, with the exception for the linear regression, the Bayesian inference of (even basic) GLMs cannot be performed without approximations due to the non-existence of convenient conjugate prior distributions. This issue is particularly pronounced in sequential (online) modeling with limited computational performance and time, making the otherwise dominant MCMC-based inference of GLMs [11] inapplicable. Second, the ZIP model is a mixture model, which increases the problem complexity. For the *basic Poisson regression model*, the method proposed in [13] provides the solution to the first difficulty.

3.1 Estimation of ϕ_t

We assume that at each time instant $t = 1, 2, \ldots$ the measurement y_t is generated by a single active component $k_t \in \{\mathcal{D}, \mathcal{P}\}$. With probability ϕ_t it is the Dirac component, and with probability $1-\phi_t$ it is the Poisson component, cf. Formula (5). The joint density of (y_t, k_t) thus reads

$$\begin{aligned} f(y_t,k_t|x_t,\beta_t,\phi_t) = {} & [\phi_t\,\delta_0]^{\mathbb{1}_{\mathcal{D}}(k_t)} \\ & \times \left[(1-\phi_t) f_{\mathcal{P}}(y_t|x_t,\beta_t)\right]^{\mathbb{1}_{\mathcal{P}}(k_t)}, \end{aligned} \tag{9}$$

where $\mathbb{1}_{\mathcal{D}}(k_t)$ and $\mathbb{1}_{\mathcal{P}}(k_t)$ are the indicators of the Dirac and Poisson component, respectively. They are equal to 1 if the component is active at time t, and 0 otherwise. The sequential Bayesian estimation of regression coefficients β_t and component probabilities ϕ_t relies on the joint prior density

$$\begin{aligned}\pi(\beta_t, \phi_t | X_{t-1}, Y_{t-1}, K_{t-1}) &= \pi(\beta_t | X_{t-1}, Y_{t-1}, K_{t-1}) \\ &\quad \times \pi(\phi_t | X_{t-1}, Y_{t-1}, K_{t-1}),\end{aligned} \tag{10}$$

where $X_{t-1} = \{x_0, \ldots, x_{t-1}\}$, $Y_{t-1} = \{y_0, \ldots, y_{t-1}\}$, and $K_{t-1} = \{k_0, \ldots, k_{t-1}\}$ symbolize the past regressors, measurements, and active components. The values x_0, y_0, and k_0 stand for the initial knowledge.

Recall, that the component probability $\phi_t \in [0, 1]$. This advocates the use of the beta distribution as the convenient prior distribution for its estimation. We denote its hyperparameters $\kappa_{\mathcal{D},t-1}$ and $\kappa_{\mathcal{P},t-1}$. The density then has the form

$$\pi(\phi_t | X_{t-1}, Y_{t-1}, K_{t-1}) \propto \phi_t^{\kappa_{\mathcal{D},t-1}-1}(1-\phi_t)^{\kappa_{\mathcal{P},t-1}-1}, \tag{11}$$

where $\propto$ stands for equality up to the normalizing constant. The prior distribution (10) for β_t and ϕ_t is sequentially updated by new data y_t and x_t by virtue of the Bayes' theorem,

$$\begin{aligned}&\pi(\beta_t, \phi_t | X_t, Y_t, K_t) \\ &\propto p(y_t, k_t | x_t, \beta_t, \phi_t)\, \pi(\beta_t, \phi_t | X_{t-1}, Y_{t-1}, K_{t-1}) \\ &\propto [\phi_t\, \delta_0]^{\mathbb{1}_{\mathcal{D}}(k_t)} \cdot [(1-\phi_t) f_{\mathcal{P}}(y_t | x_t, \beta_t)]^{\mathbb{1}_{\mathcal{P}}(k_t)} \\ &\times \phi_t^{\kappa_{\mathcal{D},t-1}-1}(1-\phi_t)^{\kappa_{\mathcal{P},t-1}-1} \\ &\times \pi(\beta_t | X_{t-1}, Y_{t-1}, K_{t-1}).\end{aligned} \tag{12}$$

A careful inspection of (12) reveals that there are two separate updates. First, the values $\mathbb{1}_{\mathcal{D}}(k_t)$ and $\mathbb{1}_{\mathcal{P}}(k_t)$ enter the beta distribution hyperparameters $\kappa_{\mathcal{D},t-1}$ and $\kappa_{\mathcal{P},t-1}$. Second, the pair x_t and y_t updates the prior distribution for β_t.

Since the active components indicators are not available, we take advantage of the quasi-Bayesian procedure [15] and replace them by their point estimates. For the moment, assume both ϕ_t and β_t constant – their time-varying nature will be solved in Sec. 3.3. The *prior* expected value of ϕ_t from the beta distribution then reads

$$\mathbb{E}[\phi_t | X_{t-1}, Y_{t-1}, K_{t-1}] = \frac{\kappa_{\mathcal{D},t-1}}{\kappa_{\mathcal{D},t-1} + \kappa_{\mathcal{P},t-1}}. \tag{13}$$

Now, using the Poisson component prior predictive likelihood $f_{\mathcal{P}}(y_t | x_t, X_{t-1}, Y_{t-1}, K_{t-1})$ we have

$$\begin{aligned}\widehat{\mathbb{1}}_{\mathcal{P}}(k_t) &\propto (1 - \mathbb{E}[\phi_t | X_{t-1}, Y_{t-1}, K_{t-1}]) \\ &\quad \times f_{\mathcal{P}}(y_t | x_t, X_{t-1}, Y_{t-1}, K_{t-1}),\end{aligned} \tag{14}$$

$$\widehat{\mathbb{1}}_{\mathcal{D}}(k_t) = 1 - \widehat{\mathbb{1}}_{\mathcal{P}}(k_t), \tag{15}$$

where

$$\begin{aligned}&f_{\mathcal{P}}(y_t | x_t, X_{t-1}, Y_{t-1}, K_{t-1}) \\ &= \int f_{\mathcal{P}}(y_t | x_t, \beta_t)\, \pi(\beta_t | X_{t-1}, Y_{t-1}, K_{t-1}) d\beta_t.\end{aligned} \tag{16}$$

This prior predictive likelihood is not analytically tractable, but since the Dirac distribution is concentrated at a single point, we can proceed with the plug-in principle and substitute the *prior* point estimate $\hat{\beta}_t$ directly into the Poisson model,

$$f_{\mathcal{P}}(y_t|x_t, X_{t-1}, Y_{t-1}, K_{t-1}) \approx f_{\mathcal{P}}(y_t|x_t, \hat{\beta}_t). \tag{17}$$

The same process applies to the Dirac component where the situation is trivial as it covers only the cases $y_t = 0$. From (12), (14) and (15),

$$\kappa_{\mathcal{D},t} = \kappa_{\mathcal{D},t-1} + \widehat{\mathbb{1}}_{\mathcal{D}}(k_t) \text{ and } \kappa_{\mathcal{P},t} = \kappa_{\mathcal{P},t-1} + \widehat{\mathbb{1}}_{\mathcal{P}}(k_t). \tag{18}$$

3.2 Approximate estimation of β_t

From (12) and (14), the posterior distribution of the regression coefficients β_t is given by the weighted Bayesian update

$$\begin{aligned}\pi(\beta_t|X_t, Y_t, K_t) \propto [f_{\mathcal{P}}(y_t|x_t, \beta_t)]^{\widehat{\mathbb{1}}_{\mathcal{P}}(k_t)} \\ \times \pi(\beta_t|X_{t-1}, Y_{t-1}, K_{t-1}).\end{aligned} \tag{19}$$

The Poisson component $f_{\mathcal{P}}(y_t|x_t, \beta_t)$ defined in (6) does not admit any convenient conjugate prior. However, it can be written [16]

$$f_{\mathcal{P}}(y_t|x_t, \beta_t) = \frac{\exp(\beta_t^\intercal x_t(y_t + c) - \exp(\beta_t^\intercal x_t))}{y_t!} \frac{1}{\exp(c\beta_t^\intercal x_t)} \tag{20}$$

where $c > 0$. It was shown in [17] that the densities proportional to the first fraction can be approximated by the Gaussian densities. In our paper [13] we modified this approximation (first used for static Poisson regression in [18]) by a linear regression-based calibration that improves the approximation quality for low counts. The results corresponds to $\mathcal{N}(\mu_t, \sigma_t^2)$ where

$$\begin{aligned}\mu_t &= \log(y_t + c) - \frac{0.5574}{y_t + c}, \\ \sigma_t &= \frac{1}{\sqrt{y_t + c}} + \frac{0.0724}{y_t + c} + \frac{0.2121}{(y_t + c)^2}.\end{aligned} \tag{21}$$

The role of c becomes evident from (21): while zero measurements $y_t = 0$ are rare in the 'pure' Poisson regression presented in [13] and can be safely discarded, the ZIP model admits excessive zeros that should be appropriately processed. This stabilization idea is due to Chan and Vasconcelos [16] who also suggest that the value $c = 1$ yields best results.

Now, if we use a convenient distribution $\mathcal{N}(b_{t-1}, P_{t-1})$ with $b_{t-1} \in \mathbb{R}^n$ and $P_{t-1} \in \mathbb{R}^{n\times n}$ as the prior for β_t, then the update (19) has the form

$$\begin{aligned}&\pi(\beta_t|X_t, Y_t, K_t) \\ &\propto \exp\left(-\frac{1}{2}(\beta_t^\intercal x_t - \mu_t)^2\sigma_t^{-2} - c\beta_t^\intercal x_t\right)^{\widehat{\mathbb{1}}_{\mathcal{P}}(k_t)} \\ &\times \exp\left(-\frac{1}{2}(\beta_t - b_{t-1})^\intercal P_{t-1}^{-1}(\beta_t - b_{t-1})\right) \\ &= \exp\left(-\frac{1}{2}\begin{bmatrix}-1\\ \beta_t\end{bmatrix}^\intercal \widehat{\mathbb{1}}_{\mathcal{P}}(k_t)T(x_t, y_t)\begin{bmatrix}-1\\ \beta_t\end{bmatrix}\right) \\ &\times \exp\left(-\frac{1}{2}\begin{bmatrix}-1\\ \beta_t\end{bmatrix}^\intercal \begin{bmatrix}b_{t-1}^\intercal\\ I\end{bmatrix} P_{t-1}^{-1}\begin{bmatrix}b_{t-1}^\intercal\\ I\end{bmatrix}^\intercal \begin{bmatrix}-1\\ \beta_t\end{bmatrix}\right),\end{aligned} \tag{22}$$

where the $T(x_t, y_t)$ is the sufficient statistic

$$T(x_t, y_t) = \sigma_t^{-2} \begin{bmatrix} \mu_t^2 & (\mu_t - c\sigma_t^2)x_t^\intercal \\ x_t(\mu_t - c\sigma_t^2) & x_t x_t^\intercal \end{bmatrix}. \tag{23}$$

From the multiplication of the last two rows in (22) it follows that the weighted Bayesian update approximately results in $\mathcal{N}(b_t, P_t)$ where

$$\begin{aligned} P_t &= \left(P_{t-1}^{-1} + \widehat{\mathbb{1}}_{\mathcal{P}}(k_t)\sigma_t^{-2} x_t x_t^\intercal\right)^{-1}, \\ b_t &= P_t\left(P_{t-1}^{-1} b_{t-1} + \widehat{\mathbb{1}}_{\mathcal{P}}(k_t)\sigma_t^{-2} x_t(\mu_t - c\sigma_t^2)\right). \end{aligned} \tag{24}$$

A similar result for the static Poisson regression without calibration is derived in [16].

3.3 Time-varying parameters β_t and ϕ_t

In a majority of real-world cases, the model parameters are not constant. If they vary sufficiently slowly, the idea of gradual discounting of older data from the last posterior distribution from time $t-1$ by exponential forgetting can be used [19, 20]. For β_t it amounts to exponentiation by $\alpha_\beta \in [0, 1]$,

$$\pi(\beta_t | X_{t-1}, Y_{t-1}, K_{t-1}) = [\pi(\beta_{t-1} | X_{t-1}, Y_{t-1}, K_{t-1}]^{\alpha_\beta}, \tag{25}$$

which obviously inflates the prior covariance matrix,

$$P_{t-1} \leftarrow \alpha_\beta P_{t-1}. \tag{26}$$

Analogously, the evolution $\phi_{t-1} \to \phi_t$ is facilitated by exponentiation by $\alpha_\phi \in [0, 1]$,

$$\pi(\phi_t | X_{t-1}, Y_{t-1}, K_{t-1}) = [\pi(\phi_{t-1} | X_{t-1}, Y_{t-1}, K_{t-1})]^{\alpha_\phi}, \tag{27}$$

which results in the prior hyperparameters at time t

$$\kappa_{\mathcal{D},t-1} \leftarrow \alpha_\phi \kappa_{\mathcal{D},t-1} \quad \text{and} \quad \kappa_{\mathcal{P},t-1} \leftarrow \alpha_\phi \kappa_{\mathcal{P},t-1}. \tag{28}$$

The hyperparameters of the respective distributions are then updated as described in the previous sections.

4 Diffusion estimation

Now suppose that we have a network consisting of a set $\mathcal{I}$ of agents that independently observe the process $Y_t^{(i)} \sim ZIP(\exp(\beta_t^\intercal x_t^{(i)}), \phi_t)$, where β_t and ϕ_t are global, i.e., identical for all $i \in \mathcal{I}$, while the observations $y_t^{(i)}$ and the regressors $x_t^{(i)}$ are local. Next, suppose that at each time instant t, all agents may perform one mutual exchange of their posterior densities $\pi^{(i)}(\beta_t, \phi_t|\cdot)$ with all their adjacent neighbors within 1 network hop distance. That is, the ith node can access posterior densities of its closed neighborhood $\mathcal{I}^{(i)}$.

The goal of the distributed estimation is to merge the available posterior densities in a way consistent with the Bayesian information processing. Let us first focus on $\pi^{(j)}(\phi_t|\cdot)$ of neighbors $j \in \mathcal{I}^{(i)}$. They are beta densities (11), whose hyperparameters $\kappa_{\mathcal{D},t}^{(j)}$ and $\kappa_{\mathcal{P},t}^{(j)}$

accumulate the information about the expected components activity, see (18). The merging rule

$$\bar{\kappa}_{\mathcal{D},t}^{(i)} = \frac{1}{|\mathcal{I}^{(i)}|} \sum_{j \in \mathcal{I}^{(i)}} \kappa_{\mathcal{D},t}^{(j)}, \qquad \bar{\kappa}_{\mathcal{P},t}^{(i)} = \frac{1}{|\mathcal{I}^{(i)}|} \sum_{j \in \mathcal{I}^{(i)}} \kappa_{\mathcal{P},t}^{(j)}, \tag{29}$$

where $|\mathcal{I}^{(i)}|$ denotes the cardinality of $\mathcal{I}^{(i)}$ thus corresponds to an averaged – uniformly weighted – Bayesian update by the data of the neighbors $j \in \mathcal{I}^{(i)}$.

The same reasoning applies to merging of the posterior densities $\pi^{(j)}(\beta_t|\cdot)$. Since they are normal, the result is known as the covariance intersection,

$$\begin{aligned} \bar{P}_t^{(i)} &= \left(\frac{1}{|\mathcal{I}^{(i)}|} \sum_{j \in \mathcal{I}^{(i)}} P_t^{-1,(j)} \right)^{-1}, \\ \bar{b}_t^{(i)} &= \bar{P}_t^{(i)} \left(\frac{1}{|\mathcal{I}^{(i)}|} \sum_{j \in \mathcal{I}^{(i)}} P_t^{-1,(j)} b_t^{\mathsf{T},(j)} \right). \end{aligned} \tag{30}$$

The hyperparameters (29) and (30) serve as the local prior hyperparameters at the next time step. We emphasize that these results are Kullback-Leibler-optimal [21]. Due to page limits, more details and analyses are postponed to future publication.

Algorithm 1 Diffusion ZIP Regression

For each agent $i \in \mathcal{I}$ set the initial hyperparameters $\kappa_{\mathcal{D},0}^{(i)}$, $\kappa_{\mathcal{P},0}^{(i)}$ and the prior distribution $\mathcal{N}(b_0^{(i)}, P_0^{(i)})$. Set the forgetting factors α_κ, α_β. For $t = 1, 2, \ldots$ and each node $i \in \mathcal{I}$ do:
Local estimation:

1. Gather observations $x_t^{(i)}, y_t^{(i)}$.
2. Flatten the prior distributions, Eqs. (26) and (28).
3. Calculate the component membership indicators $\widehat{\mathbb{1}}_{\mathcal{D}}(k_t^{(i)})$, $\widehat{\mathbb{1}}_{\mathcal{P}}(k_t^{(i)})$, Eqs. (14) and (15).
4. Update the prior of ϕ_t, Eq. (18).
5. Update the prior of β_t, Eq. (24).

Combination:

1. Get posterior densities $\pi^{(j)}(\phi_t|\cdot)$ and $\pi^{(j)}(\beta_t|\cdot)$ of neighbors $j \in \mathcal{I}^{(i)}$.
2. Combine the posterior hyperparameters of ϕ_t, Eq. (29).
3. Combine the posterior hyperparameters of β_t, Eq. (30).

5 Illustrative examples

Two simulation examples follow. Since to the authors' best knowledge no sequential ZIP regression algorithms exist, the examples study the convergence to the true parameter values.

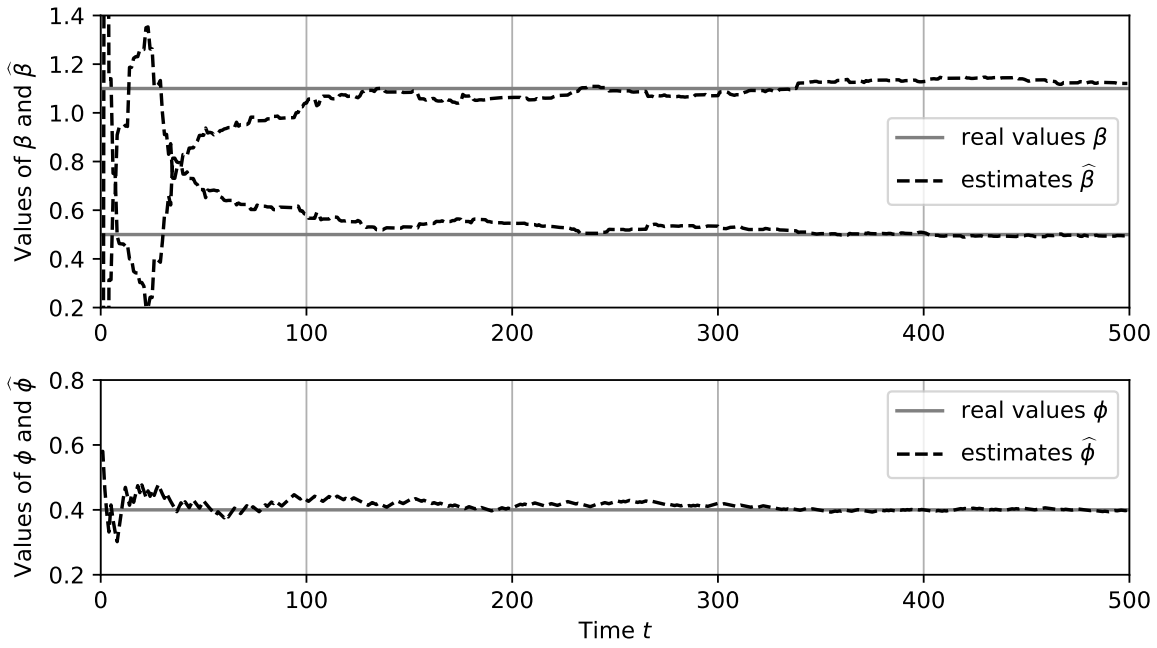

Figure 1: Example 1: The evolution of the real and estimated values of β (top) and ϕ (bottom).

The first example demonstrates the performance of the estimation method proposed in Sec. 3, i.e., without any network and collaboration. The simulated observations follow the ZIP regression process with the time-invariant parameters $\beta = [0.5, 1.1]^\intercal$, $\phi = 0.4$, and regressors $x_t \sim U\left([0.9, 2.0] \times [0.1, 1.6]\right), t = 1, \ldots, 500$. The forgetting factors $\alpha_\beta = \alpha_\phi = 1$, and the additive constant $c = 1$. The initial prior distribution for β is normal with the mean $b_0 = [0, 0]^\intercal$ and the covariance matrix $P_0 = 100 \cdot I$ where $I \in \mathbb{R}^{2\times 2}$ is the identity matrix. The prior for ϕ is beta with $\kappa_{\mathcal{D},0} = \kappa_{\mathcal{P},0} = 1$. Fig. 1 depicts the evolution of estimates of β and ϕ. Apparently, the estimates quickly converge to the true parameter values.

The second example demonstrates the performance of the proposed method in the cooperative mode. Two scenarios are compared: (i) the 'combination' scenario using Algorithm 1, and (ii) the isolated 'no combination' scenario without any collaboration among nodes. Fig. 2 depicts the randomly generated diffusion network of 50 nodes. Each node i observes independently generated outcomes of a ZIP regression process simulated with the parameters

$$\phi_t = 0.4 + 0.03 \cdot \sin\left(\pi \cdot \frac{t}{500}\right),$$
$$\beta_t = \begin{bmatrix} 0.5 + 0.1 \cdot \sin\left(\pi \cdot \frac{t}{500}\right) \\ 0.1 + 0.05 \cdot \cos\left(2\pi \cdot \frac{t}{500}\right) \\ -0.2 + 0.02 \cdot \sin\left(3\pi \cdot \frac{t}{500}\right) \end{bmatrix}, \quad t = 1, \ldots, 500,$$

and randomly generated regressors $x_t^{(i)} \sim U\left([0, 5]^3\right)$. The initial setting is identical for all $i \in \mathcal{I}$: β_t is modeled as normal with $b_0^{(i)} = [0, 0, 0]^\intercal$ and $P_0^{(i)} = 100 \cdot I$; ϕ_t is modeled as beta with $\kappa_{\mathcal{D},0} = \kappa_{\mathcal{P},0} = 1$. The additive constant $c = 1$ and the forgetting factors $\alpha_\beta = 0.98, \alpha_\phi = 0.99$. The results are averaged over 100 independent runs and all network nodes.

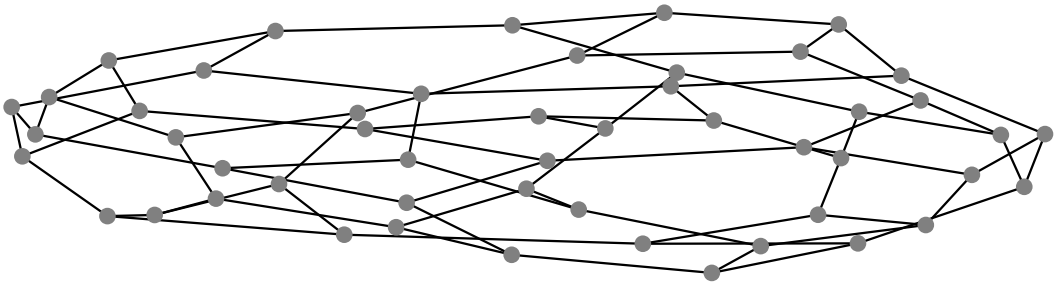

Figure 2: Example 2: Network topology used in the simulation.

Fig. 3 depicts the evolution of ARMSE – RMSE averaged over all elements of β_t, and RMSE for the estimates of ϕ_t. The collaborative estimation clearly improves the overall quality of estimates. Fig. 4 shows the evolution of raw estimates at a randomly selected node. Apparently, collaboration significantly improves estimation of time-varying parameters.

6 Conclusion

The proposed method allows reliable sequential estimation of parameters of the zero-inflated Poisson regression model from streaming data. The analytically tractable solution and the adopted communication protocol make the computational, memory, and communication requirements very low. The future work will focus on the ZIP models with multiple Poisson components, where the predictive likelihood (16) needs to be reasonably approximated. The full adapt-then-combine (ATC) strategy [2] will be adopted and properties of the whole framework will be thoroughly analyzed.

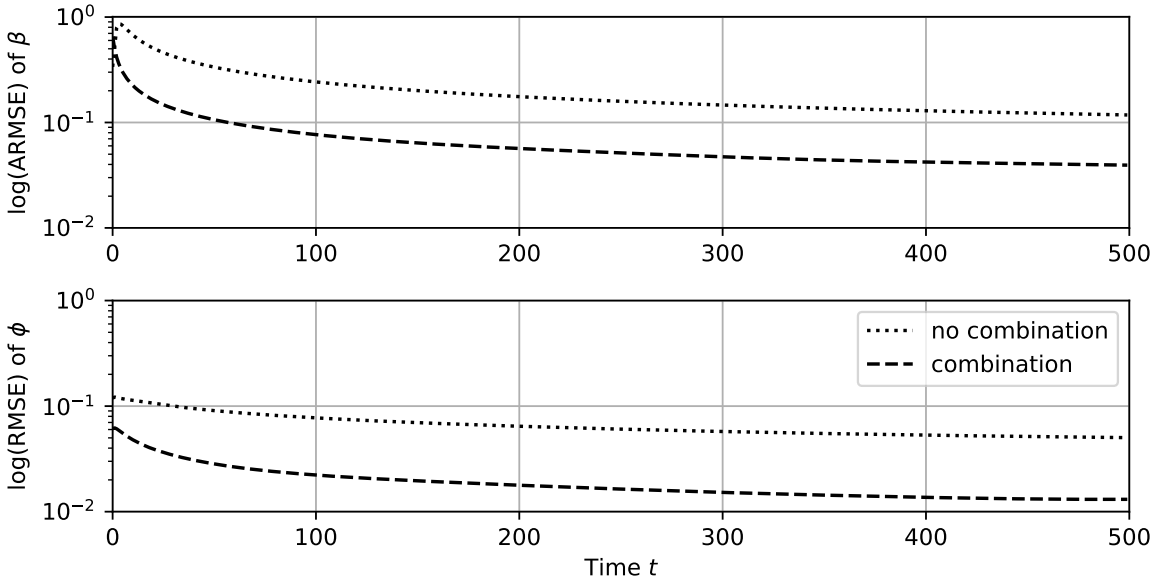

Figure 3: Example 2: Evolution of the ARMSE of the estimates of β_t (top) and RMSE of the estimates of ϕ_t (bottom) averaged over all network nodes and 100 independent experiment runs.

References

[1] A. H. Sayed, "Diffusion Adaptation over Networks," in *Academic Press Library in Signal Processing*, R. Chellapa and S. Theodoridis, Eds. Academic Press, Elsevier, May 2014, vol. 3, pp. 323–454.

[2] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends in Machine Learning*, vol. 7, no. 4-5, pp. 311–801, 2014.

[3] R. Nassif, S. Vlaski, C. Richard, J. Chen, and A. H. Sayed, "Multitask learning over graphs: An approach for distributed, streaming machine learning," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 14-–25, May 2020.

[4] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1865–1877, May 2008.

[5] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[6] J. Plata-Chaves, N. Bogdanović, and K. Berberidis, "Distributed diffusion-based LMS for node-specific adaptive parameter estimation," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3448–3460, Jul. 2015.

[7] J. Plata-Chaves, M. H. Bahari, M. Moonen, and A. Bertrand, "Unsupervised diffusion-based LMS for node-specific parameter estimation over wireless sensor networks," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Mar. 2016, pp. 4159–4163.

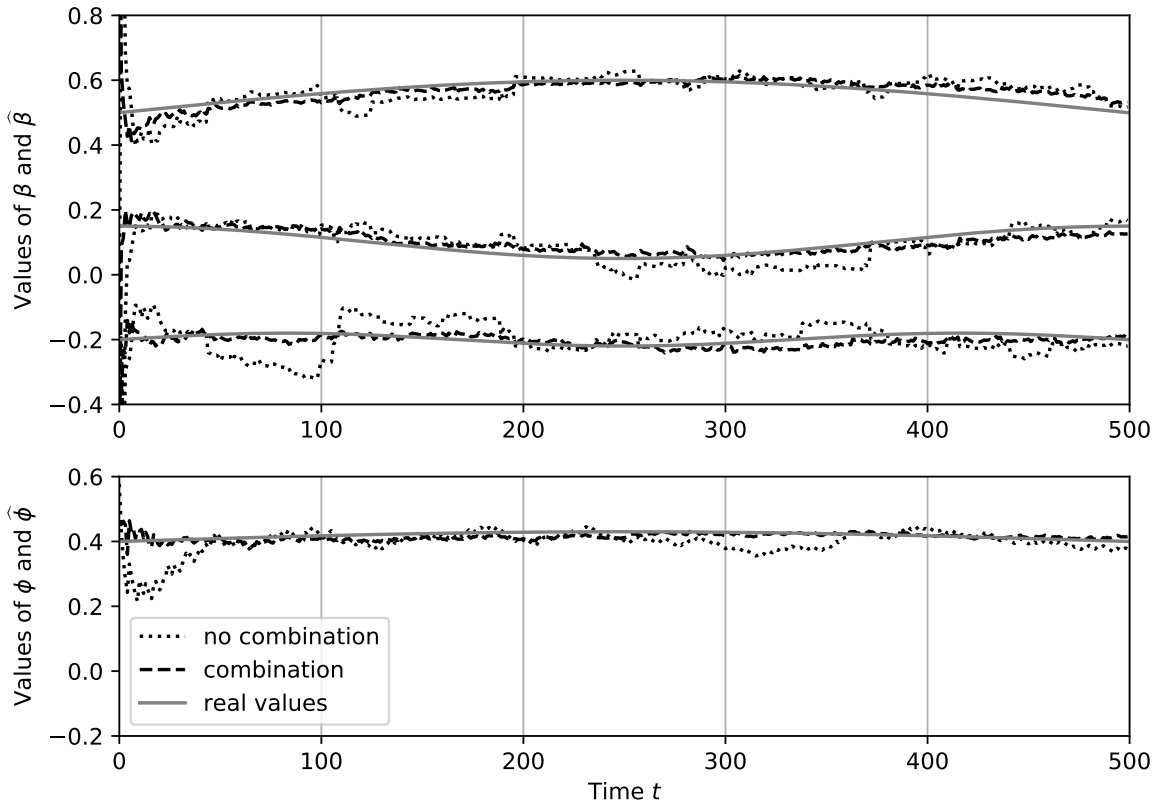

Figure 4: Example 2: Real and estimated values of β_t (top) and ϕ_t (bottom) at a randomly selected node.

[8] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 2069–2084, 2010.

[9] J. Hu, L. Xie, and C. Zhang, "Diffusion Kalman filtering based on covariance intersection," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 891–902, Feb. 2012.

[10] P. McCullagh and J. A. Nelder, *Generalized Linear Models, Second Edition*, Monographs on Statistics and Applied Probability Series. Chapman & Hall, 1989.

[11] R. H. Myers, D. C. Montgomery, G. G. Vining, and T. J. Robinson, *Generalized Linear Models*, Wiley Series in Probability and Statistics. Hoboken, NJ, USA: Wiley & Sons., Mar. 2010.

[12] D. Manolakis, N. Bosowski and V. K. Ingle, "Count time-series analysis: A signal processing serspective," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 64–81, May 2019.

[13] K. Dedecius, R. Žemlička, "Sequential Poisson Regression in Diffusion Networks," *IEEE Signal Processing Lett.*, vol. 27, pp. 625–629, 2020.

[14] D. Lambert, "Zero-inflated Poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 34, no. 1, pp. 1–14, 1992.

[15] M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma and L. Tesař, "*Optimized Bayesian Dynamic Advising: Theory and Algorithms (Advanced Information and Knowledge Processing)*," London, UK: Springer, 2006.

[16] A. B. Chan, N. Vasconcelos, "Counting people with low-level features and Bayesian regression," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2160–2177, 2011.

[17] M. S. Bartlett and D. G. Kendall, "The statistical analysis of variance-heterogeneity and the logarithmic transformation," *Supplement to the Journal of the Royal Statistical Society*, vol. 8, no. 1, pp. 128–138, 1946.

[18] G. M. El-Sayyad, "Bayesian and classical analysis of Poisson regression," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 35, no. 3, pp. 445–451, Jul. 1973.

[19] K. Dedecius, I. Nagy, and M. Kárný, "Parameter tracking with partial forgetting method," *International Journal of Adaptive Control and Signal Processing*, vol. 26, no. 1, pp. 1–12, 2012.

[20] V. Peterka, "Bayesian approach to system identification," in *Trends and Progress in System Identification*, P. Eykhoff, Ed. Oxford, U.K.: Pergamon Press, 1981, pp. 239–304.

[21] K. Dedecius and P. M. Djurić, "Sequential estimation and diffusion of information over networks: A Bayesian approach with exponential family of distributions," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1795–1809, Apr. 2017.